[BOB7기][포렌식][김성수][실습과제]

download files from website, parse info, make db and analysis

/* 개요 */

개발환경 및 설정

요구사항 및 결과

구현방법

첨부파일 설명

개발 환경 및 설정

1. 개발환경

OS : Windows 7 Ultimate

interpreter : python 3.7

2. 추가 라이브러리

가. requests

Website 를 읽는데 사용

나. BeatifulSoup

Website 에서 Crawling 하는데 사용

다. zipfile

다운로드 파일의 압축을 푸는데 사용

라. sqlite3

데이터베이스를 만드는데 사용

요구사항 및 결과

1. 요구사항

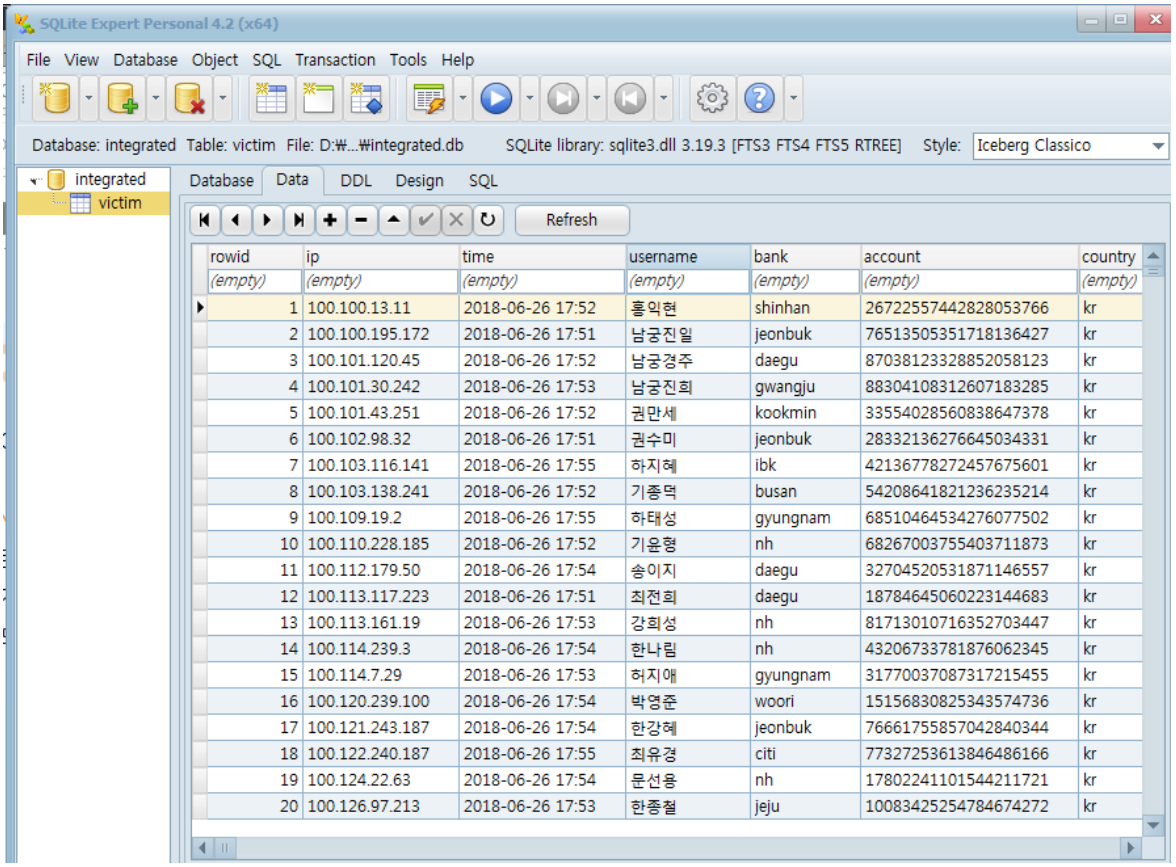
1) 피해자의 일련번호, 피해시각(서버에 피해자의 공인인증서가 업로드된 시간), 피해자의 이름, 은행명, 계좌번호, ip 주소, 피해자의 현재 소재지(국가) 를 데이터베이스(mysql)에 저장하시오

2) 은행별 유출된 공인인증서의 갯수를 계산하시오

2. 피해자 관련 정보 데이터베이스 구축(integrated.db)

제한사항1) 피해자의 현재 소재지를 ip 주소를 통해 조회하려 하였으나, kisa 에서 일정갯수를 초과하여 질의할 경우, 세션을 끊음.

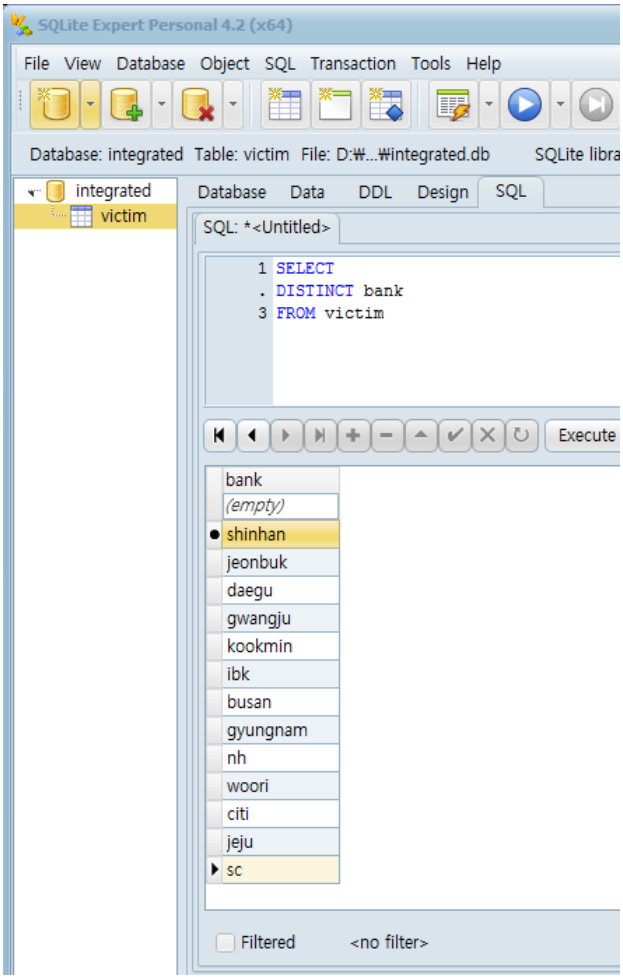
제한사항2) 파일 다운로드 과정 중, 2개가 exception 오류로 다운이 안되었으나, 오차범위 이내이므로 무시한다. (총 다운로드 된 파일 개수 : 36,832)



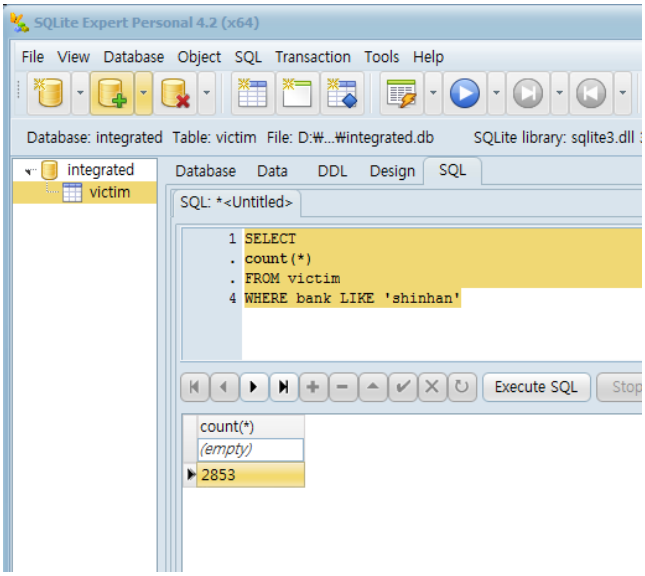
rowid	ip	time	username	bank	account	country
(empty)	(empty)	(empty)	(empty)	(empty)	(empty)	(empty)
1	100.100.13.11	2018-06-26 17:52	홍익현	shinhan	26722557442828053766	kr
2	100.100.195.172	2018-06-26 17:51	남궁진일	jeonbuk	76513505351718136427	kr
3	100.101.120.45	2018-06-26 17:52	남궁경주	daegu	87038123328852058123	kr
4	100.101.30.242	2018-06-26 17:53	남궁진희	gwangju	88304108312607183285	kr
5	100.101.43.251	2018-06-26 17:52	권만세	kookmin	33554028560838647378	kr
6	100.102.98.32	2018-06-26 17:51	권수미	jeonbuk	28332136276645034331	kr
7	100.103.116.141	2018-06-26 17:55	하지혜	ibk	42136778272457675601	kr
8	100.103.138.241	2018-06-26 17:52	기종덕	busan	54208641821236235214	kr
9	100.109.19.2	2018-06-26 17:55	하태성	gyungnam	68510464534276077502	kr
10	100.110.228.185	2018-06-26 17:52	기윤형	nh	68267003755403711873	kr
11	100.112.179.50	2018-06-26 17:54	송이지	daegu	32704520531871146557	kr
12	100.113.117.223	2018-06-26 17:51	최전희	daegu	18784645060223144683	kr
13	100.113.161.19	2018-06-26 17:53	강희성	nh	81713010716352703447	kr
14	100.114.239.3	2018-06-26 17:54	한나림	nh	43206733781876062345	kr
15	100.114.7.29	2018-06-26 17:53	허지애	gyungnam	31770037087317215455	kr
16	100.120.239.100	2018-06-26 17:54	박영준	woori	15156830825343574736	kr
17	100.121.243.187	2018-06-26 17:54	한강혜	jeonbuk	76661755857042840344	kr
18	100.122.240.187	2018-06-26 17:55	최유경	citi	77327253613846486166	kr
19	100.124.22.63	2018-06-26 17:54	문선용	nh	17802241101544211721	kr
20	100.126.97.213	2018-06-26 17:53	한종철	jeju	10083425254784674272	kr

3. 은행별 유출된 공인인증서 갯수

1) 은행종류



2) 은행별 공인인증서 갯수 구하기 예시



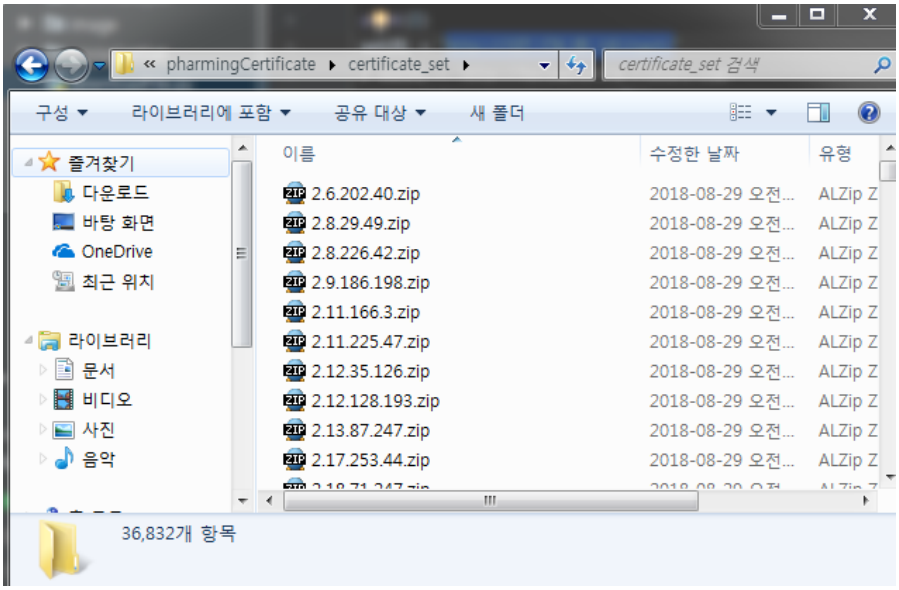
3) 은행별 탈취된 공인인증서 갯수

은행	갯수
shinhan	2853
jeonbuk	2723
daegu	2920
gwangju	2891
kookmin	2725
ibk	2734
busan	2823
gyungnam	2895
nh	2949
woori	2729
citi	2808
jeju	2879
sc	2903

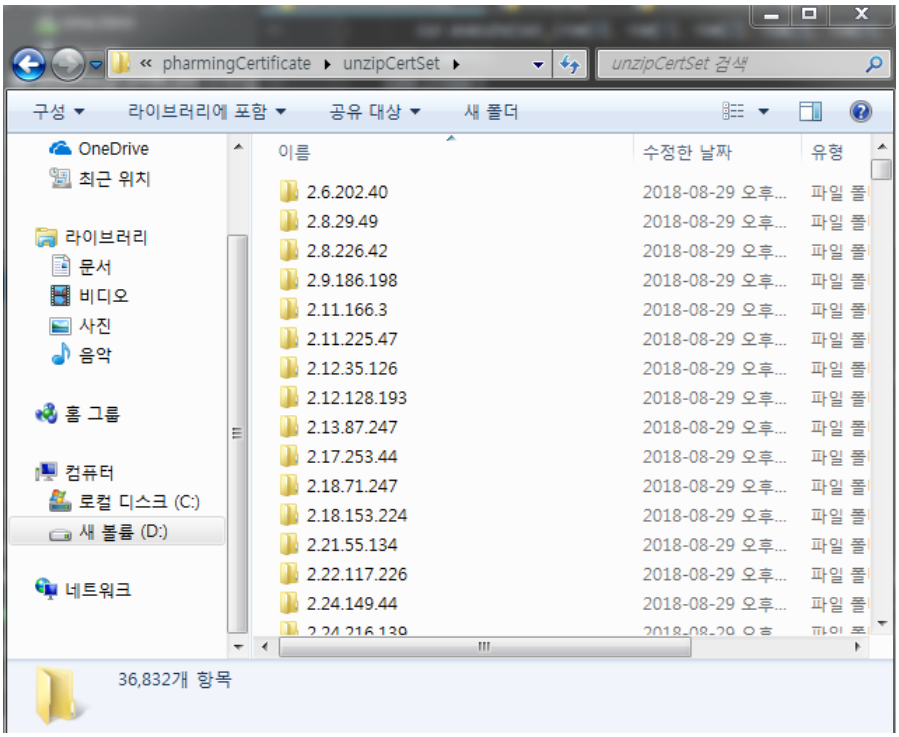
구현방법

1. downloadFile.py 를 통해, http://107.174.85.141/cert/ 에서 파일을 다운로드한다.

(프로그램을 돌린 후, 자고 일어나니 36,832/36,834 개가 다운로드 되었다. 2개는 받는 도중 에러가 나서 exception 처리 되었으며, 오차범위 이내이므로 무시한다.)



2. unzip.py 를 통해 각각의 ip 를 폴더로 하여, 그 안에 공인인증서를 압축을 푼다.



3. DB를 만들기 전에, webpage에서 text를 파싱하는 속도가 느려, downloadWebPage.py 를 통해 http://107.174.85.141/cert/ 에 해당하는 데이터를 받아서, local 에 "webpageTxt" 로 저장하였다.

4. parseAndMakeDB.py

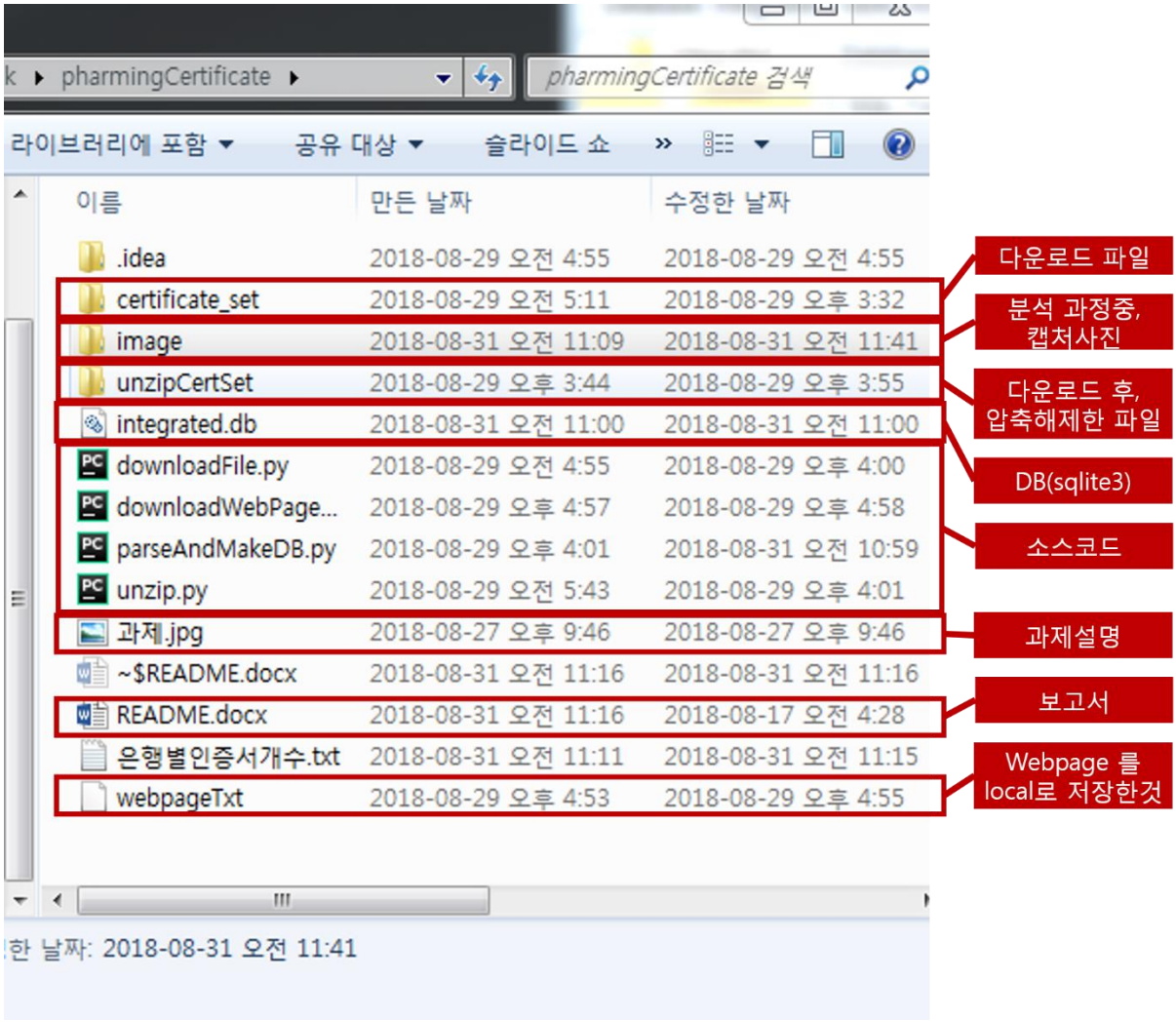
1) 공인인증서에서 [ip, username, account, bank, country] 에 해당하는 정보를 추출하여, CertDB 라는 리스트를 만든다.

2) webpageTxt에서 {ip, modified time} 에 해당하는 정보를 추출하여, WebDB 라는 딕셔너리를 만든다.(딕셔너리로 구현한 이후는, 이후에 DB 를 통합할때 탐색 속도를 빠르게 하기 위해서다.)

3) CertDB 와 WebDB 에서 ip 를 공통키로 하여, [ip, modified time, username, bank, account, country] 내용을 가진 integratedDB 라는 리스트를 만든다.

4) sqlite3 를 통해 "integrated.db" 를 만든다. 이때, 피해자의 일련번호에 해당하는 rowid 라는 컬럼은 자동으로 추가된다.

첨부파일 설명


pharmingCertificate
pharmingCertificate 검색
라이브러리에 포함
공유 대상
슬라이드 쇼
이름
만든 날짜
수정한 날짜
.idea 2018-08-29 오전 4:55 2018-08-29 오전 4:55
certificate_set 2018-08-29 오전 5:11 2018-08-29 오후 3:32
image 2018-08-31 오전 11:09 2018-08-31 오전 11:41
unzipCertSet 2018-08-29 오후 3:44 2018-08-29 오후 3:55
integrated.db 2018-08-31 오전 11:00 2018-08-31 오전 11:00
downloadFile.py 2018-08-29 오전 4:55 2018-08-29 오후 4:00
downloadWebPage... 2018-08-29 오후 4:57 2018-08-29 오후 4:58
parseAndMakeDB.py 2018-08-29 오후 4:01 2018-08-31 오전 10:59
unzip.py 2018-08-29 오전 5:43 2018-08-29 오후 4:01
과제.jpg 2018-08-27 오후 9:46 2018-08-27 오후 9:46
~$README.docx 2018-08-31 오전 11:16 2018-08-31 오전 11:16
README.docx 2018-08-31 오전 11:16 2018-08-17 오전 4:28
은행별인증서개수.txt 2018-08-31 오전 11:11 2018-08-31 오전 11:15
webpageTxt 2018-08-29 오후 4:53 2018-08-29 오후 4:55
한 날짜: 2018-08-31 오전 11:41
다운로드 파일
분석 과정중, 캡처사진
다운로드 후, 압축해제한 파일
DB(sqlite3)
소스코드
과제설명
보고서
Webpage 를 local로 저장한것